APPENDIX A

Evidence Table

Author/Year	Study Design	Demographics	Interventions Outcome Measures Instrument	Results	Methodologic Comments
Gehi, Stein, Metz, Gomes 2005	Meta-analysis of prospective studies of the predictive value of exercise-induced MTWA published from Jan 1990 to December 2004.	2608 total subjects 19 studies met inclusion/exclusion criteria. Wide range of populations included in analysis: CHF, ischemic CHF, non-ischemic CHF, post MI, athletes and healthy subjects.	MTWA used as diagnostic test Endpoints included SCD, T, VF, ICD placement, cardiac death PPV, NPV and RR computed	Presence of MTWA predicted a 4-fold higher risk of VAE. For all studies, PPV=19.3 (CI 18-21) NPV=97.2 (CI 97-98) RR=3.77 (CI 2.4-6) For CHF, PPV=25.5 (CI 23-28) NPV=93.8 (CI 92-95) RR=2.51 (CI 1.7-3.6) For post MI, PPV=6 (CI 4.5-7.4) NPV=99 (CI 99-100) RR=4.74 (CI 1.1-20.1)	Meta –analysis, no evid of publication bias or lack of heterogeneity Unable to determine the incremental prognostic value of MTWA independent of other predictors of arrhythmic events End pts of the individual studies used in summary calculations were variable Subjects primarily male Inconsistency in the exclusion of subjects using beta blockers or anti-arrhythmic meds
Gold, Bloomfield, Anderson, El-Sherif, Wilber, Groh, Estes, Greenberg,	Prospective, multi-center	313 participants, had to have NSR and capable of bicycle exercise	MTWA, SAE, and ventricular stimulator (EPS) were diagnostic tools	For MTWA, Sn=77.8% Sp=72.5% PPV=42.9% NPV=92.5%	Heterogeneous pt population Majority of VTE were nonfatal

Rosenbaum 2000		Mean age 56+/- 16, mean EF 44% +/- 18% 34% had history of CHF, including 22% with NYHA Class II symptoms, and 12% with Class III symptoms No structural heart dis in 30% of this cohort	VTE and death as endpoints Sn, Sp, PPV, NPV and RR	RR=5.7 For SAE, Sn=55.6% Sp=83.3% PPV=46.9% NPV=87.65% RR=3.8 For MTWA, w/VTE as endpoint, RR=6.1, and w/VTE or death as endpoint RR=8 For SAE, w/VTE as endpoint, RR=4.6, and w/VTE or death as endpoint RR=2.9	No powered to assess the predictors of mortality only
Hohnloser, Klingenheben, Bloomfield, Dabbous, Cohen 2003	Prospective observational study; 87 participants taken from Ikeda and colleague study, and 42 subjects taken from Klingenheben study.	129 participants Eligibility criteria included: confirmed dx of dilated cardiomyopathy, no intercurrent illnesses limiting life expectancy, sinus rhythm at initial presentation Mean age 55, 77% male 18 month follow up	Endpoints included: sudden death, cardiac arrest due to VF, or hemodynamically unstable VT or VF Diagnostic tools included: MTWA, LVEF, BRS, SAE, SDNN, IVCD, NSVT Sn, Sp, PPV, NPV, RR computed	MTWA pos in 48%, neg in 25%, indeterminate in 27% of participants Multivariate analysis revealed that MTWA was the only statistically signif predictor of arrhythmic events (Chi-square 3.67) For MTWA, Sn-87% Sp=38% PPV=22% NPV=94% RR=3.4 For SAE, Sn=47%	

				Sp=63% PPV=17% NPV=88% RR=1.4 For LVEF, Sn=80% Sp=21% PPV=15% NPV=8.6% RR=1.0	
Kitamira, Ohnishi, Okajima, Ishida, Galeano, Adachi, Yokoyama 2002	Prospective observational	104 patients with dilated cardiomyopathy (84 males) with mean age 52 24 pts Group A 22 pts Group B	Endpoints include SCD, SVT, VF Diagnostic tools included: MTWA, LVEF, SAE, LVDd	Of the 104 patients, 46 were pos for MTWA, 37 were neg, 21 were indeterminate 83 of 104 were reported at follow up For Group A MTWA pos, there were 9 cardiac events; for Group B MTWA pos, there were 2 cardiac events; for indeterminate there was 1 cardiac event Determination of OHR in combination w/MTWA can identify the high risk subgroup among the 83 pts with dilated cardiomyopathy. Cox hazard analysis revealed that MTWA with an OHR ≤ 100 bpm, and LVEF were independent predictors	Results are based on 83 pts (20% of pts lost to follow up) Low number of arrhythmic events could skew data Cut-off for OHR ≤ 100 bpm needs to be validated Sn, Sp, PPV, NPV not used.

				of arrhythmic events.	
Adachi, Ohnishi, Yokoyama 2001	Prospective observational	82 consecutive pts, mean age 53, 81% male 10 participants in Group A (high risk) 54 participants in Group B (low risk)	Endpts include SCD, SVT, VF Diagnostic tools included MTWA, LVEF, SAE, LVDd, NSVT, QTd	Participants in Group A had more arrhythmic events that those in Group B (90% v 39%) Combination of LVEF ≤ 35% and MTWA were the only statistically signif independent predictors of arrhythmic risk For MTWA, Sn=90% Sp=61% PPV=30% NPV=97% RR=10.2 For SAE, Sn=40% Sp=80% PPV=27% NPV=88% RR=2.2 For LVEF, Sn=70% Sp=80% PPV=39% NPV=93% RR=6	
Momiyama, Hartikainen, Nagayoshi, Albrecht, Kautzner, Saumarez, McKenna, Camm	14 pts with HCM were compared to 9 controls Risk stratification for VTEs made before the study, based on	7 high risk (mean age 32), 7 low risk (mean age 31) and 9 control (mean age 34) Approx equal	MTWA used as diagnostic tool Endpoints included VTEs	Alternans voltage higher in the high risk compared to low risk and control groups (2.8 v 0.6 v 0.3 respectively)	Small sample size Sn, Sp, PPV, NPV not used

1997	adverse fam hx, detection of VT on ambulatory EKG, and the findings of paced ventriculograms	males:females		In the high risk group the median alternans ratio was also higher that the low risk and controls (3.9 v 0.6 v 0.3 respectively) Of the 7 high risk pts, 5 (71%) had signof alternans	
Ikeda, Sakata, Takami, Kondo, Tezuka, Nakae, Noro, Enjoji, Abe, Sugi 2000	Prospective with consecutive pts	102 pts adm to CCU between Feb 1997 and Nov 1998 with MI dx Mean age 61.6	Late potentials analyzed using SAE, MTWA, and LVEF were used as measures Arrhythmic events (spont vent arrhythmias, sustained ventricular arrhythmias, non-sustained ventricular arrhythmias, and ventricular fibrillation Diag measures inclue Sn, Sp, PPV, NPV, hazard ratio (RH)	MTWA present in 49% of pts, while LP and reduced EF were present in 21% and 27% of pts respectively. During the followup period, VTE occurred in 15% of pts. Event rates were signif higher in pts w/MTWA, LP, or decreased EF. For MTWA, Sn=93% Sp=59% PPV=28% NPV=98% RH=16.8 For LP, Sn=53% Sp=85% PPV=38% NPV=91% RH=5.7 For EF, Sn=60% Sp=78% PPV=32%	Small sample size

				NPV = 92% RH=4.7	
Ikeda, Saito, Tanno, Shiizu, Watanabe, Ohnishi, Kasamaki, Ozawa 2002	Prospective with consecutive enrollment	850 initially enrolled, but only 834 included in study Mean age 70	Endpoints include SCD, resuscitated VF, sustained VF Outcome measures include MTWA, EF, LP Diag measures include Sn, Sp, PPV, NPV, RH	MTWA positive in 36%, neg in 52%, indeterminate in 12%. EF abnl in 18%, and LP was pos in 18%. For MTWA, Sn=92% Sp=61% PPV=7% NPV=99% RH=11.4 For EF, Sn=56% Sp=83% PPV=9% NPV=98% RH=6.6 For MTWA/EF, Sn=52% Sp=92% PPV=8% NPV=98% RH=11.9 For LP, Sn=50% Sp=84% PPV=10% NPV=98% RH=5.2	Heart rate variability was not included
Bloomfield, Steinman, Namerow, Parides,	Epidemiological study with samples from 11 clinical centers in the	549 subjects, had to be 18 or older with LVEF ≤ 40% and no	All-cause mortality endpoint	For all MADIT II-like pts, actuarial 2-year mortality was 13.2%. Based on 2-	Accuracy measures such as Sn, SP, PPV, NPV not used

Dividenko, Russo, Tang, Bigger 2004	US	prior hx of arrhythmic event 177 had MADIT-II-like characteristics Patients with atrial fib or flutter were excluded	MTWA and QRS duration were measures	yr actuarial mortality data, pts w/abnl MTWA (17.8%) had a higher mortality rate than pts w/nl MTWA. For MTWA, actuarial mortality was 17.8% for abnl test, 3.8% for nl test, hazard ratio 4.8; 32.2% were classified as low risk. False neg rate 3.5% For QRS duration, actuarial mortality was 15.9% for abnl test, 12% for nl test. , hazard ratio 1.5; 68.2% of pts were classified as low risk. False neg rate 10.2%	
Cohen 2003	Review	9 studies included Study size ranged from 82-834. Follow up period ranfged from 13-72 mos. Population suffered from variety of conditions: MI, CHF, dilated cardiomyopathy, referred for electrophysiologic studies.	VTE endpoints MTWA was the only outcome measure mentioned RR was the measure of association measured	RR ranged btwn 1.4 and 16.8 According to the aauthor MTWA was shown to be effective across a number of pt populations	No inclusion criteria included in selecting the articles to review. Sn, SP, PPV, NPV not reported
Hohnloser, Ikeda, Bloomfield, Dabbous, Cohen	Subgroup analysis of 2 prior studies (Ikeda et al 2002 and Klingenheben	129 pts, all w/prior MI and EF ≤ 30%; 112 males, mean age 63,	SCD was endpt MTWA was the only	Mortality rate among pts w/neg MTWA was 42% lower than among the	Sn, SP, PPV, NPV not included

2003	et al 2000) which evaluated the used of MTWA in MADIT II type pts	mean duration of follow up 16 mos	outcome measure	non-neg pts. No SCD in pts w/neg MTWA test, but 10 pts pos for MTWA and 2 pts indeterminate for MTWA had cardiac events	
Grimm, Christ, Bach, Muller, Maisch 2003	Prospective observational, with enrollment between March 1996 and June 2001 (MACAS study)	343 participants, including 263 w/sinus rhythm and 80 with a-fib at study entry. Follow up for 52 mos Men and women between 16 and 70 w/ICDs and LV end-diast diam 56 mm by echo. Exclusions include hx of NYHA Class IV, hx of sustained VT or VF, CAD (50% stenosis by angiogram), hx of MI, HBP	VTEs and SCDs were the endpoints Diag tests include: LVEF and size, QTc dispersion, SAE, arrhythmias on Holter, heart rate variability, baroflex sensitivity, MTWA	46 pts (13%) experienced sustained VT, VF, or SCD. On multivariate analysis, LVEF was the only signif arrhythmia risk predictor in pts w/sinus rhythm (RR of 2.3 per 10% decrease in EF) On multivariate analysis, LVEF was also the only signif predictor in pts with heart transplant (RR of 2.51 per 10% decrease in EF) MTWA did not seem to be helpful for arrhythmia risk stratification	Pts w/ NYHA Class IV were excluded (other studies included Class III and Class IV pts) Pts w/CAD were excluded (most other studies included pts w/MI) Sn. Sp, PPV, NPV not done

Appendix B: General Methodological Principles of Study Design

When making national coverage determinations, CMS evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service falling within a benefit category is reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member. The critical appraisal of the evidence enables us to determine whether: 1) the specific assessment questions can be answered conclusively; and 2) the intervention will improve net health outcomes for patients. An improved net health outcome is one of several considerations in determining whether an item or service is reasonable and necessary.

CMS divides the assessment of clinical evidence into three stages: 1) the quality of the individual studies; 2) the relevance of findings from individual studies to the Medicare population; and 3) overarching conclusions that can be drawn from the body of the evidence on the direction and magnitude of the intervention's risks and benefits.

The issues presented here represent a broad discussion of the issues we consider when reviewing clinical evidence. However, it should be noted that each coverage determination has unique methodological aspects.

1. Assessing Individual Studies

Methodologists have developed criteria to determine weaknesses and strengths of clinical research. Strength of evidence generally refers to: 1) the scientific validity underlying study findings regarding causal relationships between health care interventions and health outcomes; and 2) the reduction of bias. In general, some of the methodological attributes associated with stronger evidence include those listed below:

- Use of randomization (allocation of patients to either intervention or control group) in order to minimize bias.
- Use of contemporaneous control groups (rather than historical controls) in order to ensure comparability between the intervention and control groups.
- Prospective (rather than retrospective) studies to ensure a more thorough and systematical assessment of factors related to outcomes.
- Larger sample sizes in studies to help ensure adequate numbers of patients are enrolled to demonstrate both statistically significant as well as clinically significant outcomes that can be extrapolated to the Medicare population. Sample size should be large enough to make chance an unlikely explanation for what was found.
- Masking (blinding) to ensure patients and investigators do not know to which group patients were assigned (intervention or control). This is important especially in subjective outcomes, such as pain or quality of life, where enthusiasm and psychological factors may lead to an improved perceived outcome by either the patient or assessor.

Regardless of whether the design of a study is a randomized controlled trial, a non-randomized controlled trial, a cohort study or a case-control study, the primary criterion for methodological strength or quality is the extent to which differences between intervention and control groups can

be attributed to the intervention studied. This is known as internal validity. Various types of bias can undermine internal validity. These include:

- Different characteristics between patients participating and those theoretically eligible for study but not participating (selection bias)
- Co-interventions or provision of care apart from the intervention under evaluation (confounding)
- Differential assessment of outcome (detection bias)
- Occurrence and reporting of patients who do not complete the study (attrition bias)

In principle, rankings of research design have been based on the ability of each study design category to minimize these biases. A randomized controlled trial minimizes systematic bias (in theory) by selecting a sample of participants from a particular population and allocating them randomly to the intervention and control groups. Thus, randomized controlled studies have been typically assigned the greatest strength, followed by non-randomized clinical trials and controlled observational studies. The following is a representative list of study designs (some of which have alternative names) ranked from most to least methodologically rigorous in their potential ability to minimize systematic bias:

- Randomized controlled trials
- Non-randomized controlled trials
- Prospective cohort studies
- Retrospective case control studies
- Cross-sectional studies
- Surveillance studies (e.g., using registries or surveys)
- Consecutive case series
- Single case reports

When there are merely associations but not causal relationships between a study's variables and outcomes, it is important not to draw causal inferences. Confounding refers to independent variables that systematically vary with the causal variable. This distorts measurement of the outcome of interest because its effect size is mixed with the effects of other extraneous factors. For observational, and in some cases randomized controlled trials, the method in which confounding factors are handled (either through stratification or appropriate statistical modeling) are of particular concern. For example, in order to interpret and generalize conclusions to our population of Medicare patients, it may be necessary for studies to match or stratify their intervention and control groups by patient age or co-morbidities.

Methodological strength is, therefore, a multidimensional concept that relates to the design, implementation, and analysis of a clinical study. In addition, thorough documentation of the conduct of the research, particularly study's selection criteria, rate of attrition and process for data collection, is essential for CMS to adequately assess the evidence.

2. Generalizability of Clinical Evidence to the Medicare Population

The applicability of the results of a study to other populations, settings, treatment regimens, and outcomes assessed is known as external validity. Even well-designed and well-conducted trials may not supply the evidence needed if the results of a study are not applicable to the Medicare population. Evidence that provides accurate information about a population or setting not well represented in the Medicare program would be considered but would suffer from limited generalizability.

The extent to which the results of a trial are applicable to other circumstances is often a matter of judgment that depends on specific study characteristics, primarily the patient population studied (age, sex, severity of disease, and presence of co-morbidities) and the care setting (primary to tertiary level of care, as well as the experience and specialization of the care provider). Additional relevant variables are treatment regimens (dosage, timing, and route of administration), co-interventions or concomitant therapies, and type of outcome and length of follow-up.

The level of care and the experience of the providers in the study are other crucial elements in assessing a study's external validity. Trial participants in an academic medical center may receive more or different attention than is typically available in non-tertiary settings. For example, an investigator's lengthy and detailed explanations of the potential benefits of the intervention and/or the use of new equipment provided to the academic center by the study sponsor may raise doubts about the applicability of study findings to community practice.

Given the evidence available in the research literature, some degree of generalization about an intervention's potential benefits and harms is invariably required in making coverage decisions for the Medicare population. Conditions that assist us in making reasonable generalizations are biologic plausibility, similarities between the populations studied and Medicare patients (age, sex, ethnicity and clinical presentation), and similarities of the intervention studied to those that would be routinely available in community practice.

A study's selected outcomes are an important consideration in generalizing available clinical evidence to Medicare coverage determinations because one of the goals of our determination process is to assess net health outcomes. We are interested in the results of changed patient management not just altered management. These outcomes include resultant risks and benefits such as increased or decreased morbidity and mortality. In order to make this determination, it is often necessary to evaluate whether the strength of the evidence is adequate to draw conclusions about the direction and magnitude of each individual outcome relevant to the intervention under study. In addition, it is important that an intervention's benefits are clinically significant and durable, rather than marginal or short-lived.

If key health outcomes have not been studied or the direction of clinical effect is inconclusive, we may also evaluate the strength and adequacy of indirect evidence linking intermediate or surrogate outcomes to our outcomes of interest.

3. Assessing the Relative Magnitude of Risks and Benefits

Generally, an intervention is not reasonable and necessary if its risks outweigh its benefits. Net health outcomes are one of several considerations in determining whether an item or service is reasonable and necessary. For most determinations, CMS evaluates whether reported benefits translate into improved net health outcomes. CMS places greater emphasis on health outcomes actually experienced by patients, such as quality of life, functional status, duration of disability, morbidity, and mortality, and less emphasis on outcomes that patients do not directly experience, such as intermediate outcomes, surrogate outcomes, and laboratory or radiographic responses. The direction, magnitude, and consistency of the risks and benefits across studies are also important considerations. Based on the analysis of the strength of the evidence, CMS assesses the relative magnitude of an intervention or technology's benefits and risk of harm to Medicare beneficiaries.